

Multi-camera infrared thermography for infant respiration monitoring

ILDE LORATO,^{1,*}  SANDER STUIJK,¹ MOHAMMED MEFTAH,²
DEEDEE KOMMERS,^{3,4} PETER ANDRIESEN,^{3,4} CAROLA VAN
PUL,^{4,5} AND GERARD DE HAAN¹

¹Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands

²Department of Family Care Solutions, Philips Research, Eindhoven, The Netherlands

³Department of Neonatology, Maxima Medical Centre, Veldhoven, The Netherlands

⁴Department of Applied Physics, Eindhoven University of Technology, Eindhoven, The Netherlands

⁵Department of Clinical Physics, Maxima Medical Centre, Veldhoven, The Netherlands

*i.r.lorato@tue.nl

Abstract: Respiration is monitored in neonatal wards using chest impedance (CI), which is obtrusive and can cause skin damage to the infants. Therefore, unobtrusive solutions based on infrared thermography are being investigated. This work proposes an algorithm to merge multiple thermal camera views and automatically detect the pixels containing respiration motion or flow using three features. The method was tested on 152 minutes of recordings acquired on seven infants. We performed a comparison with the CI respiration rate yielding a mean absolute error equal to 2.07 breaths/min. Merging the three features resulted in reducing the dependency on the window size typical of spectrum-based features.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Premature infants are cared for in Neonatal Intensive Care Units (NICUs) where their vital signs need to be continuously monitored to detect critical events. To measure these biosignals, like Electrocardiogram (ECG), respiration, and oxygen saturation, many electrodes and sensors are applied on the infants' sensitive skin causing discomfort and, in some cases, also skin damage [1]. Respiratory frequency is usually monitored since respiratory instability and apneas (cessations of breathing) can be common in preterm infants and in term newborns with respiratory diseases [2], and may require immediate action of the caregiver. The gold standard to monitor respiratory frequency in non-ventilated patients in NICUs is the Chest Impedance (CI) method, which can be measured using the electrodes already applied for the ECG. However, it still suffers from motion artifacts and it is not reliable for neonatal apnea detection [3].

Unobtrusive or non-contact solutions to monitor respiration are being investigated for clinical environments. The two main observable effects of respiration are respiratory motion and respiratory flow. Respiratory motion can be monitored unobtrusively using different solutions as RGB/Near-Infrared (NIR) cameras [4,5], radar-based solutions [6,7], pressure mats [8,9], or thermal cameras [10,11]. Of these, thermal cameras are able to also deliver flow information unobtrusively [12,13].

Respiratory monitoring is often used to detect apneas, which can be classified in three main categories: central, obstructive, and mixed. In central apneas no stimulus for breathing is given by the brain and therefore, both respiratory motion and respiratory flow are absent, while, in the other two categories some sort of respiratory motion, i. e. respiratory effort, is present [14]. Therefore, when aiming at apnea detection and classification, technologies that can detect both flow and motion are more appealing, like infrared thermography.

Previous studies based on thermal cameras in a NICU environment used facial landmarks detection to locate the nasal area by exploiting for example the medial canthus region, one of the

warmest areas of the face [15,16]. This is however quite complex in infants' thermal recordings and requires high resolution thermal images, or proximity of the camera to the nasal area [17], or a combination of thermal and RGB cameras [18]. Therefore, approaches aiming at automatic Region Of Interest (ROI) detection in thermal imaging are being developed [19,20].

Pereira et al. [19] designed an algorithm to automatically select the ROI containing respiration signals in infants. This algorithm relies on a large resolution reduction which allows to drastically reduce the noise in the thermal recordings. Afterwards, a Signal Quality Index (SQI) is calculated for each ROI based on the spectrum and an empirical threshold is used to decide which ROIs potentially contain respiration signals. Such an approach is very promising as it allows to obtain a respiration signal even when the nose is not clearly visible in infant thermal recordings. Our previous work [20] proposed an alternative approach tested on a thermopile array, which is a very low-cost device that delivers low-resolution thermal images. The algorithm proposed for the automatic detection of the ROI was based on the height of the normalized spectrum peak and required no empirical thresholds to be defined. While the first work was developed for a high-resolution thermal camera and was tested also in infants, the second one was developed for an extremely low-resolution thermopile array but was tested only for adults in ideal conditions. However, both approaches propose the use of features for the localization of the pixels containing respiration based solely on the spectrum shape, this may be complex to generalize as different spectrum characteristics can affect the result. For example, the frequency resolution of the spectrum affects how the energy is distributed and if short windows are used, it may result in noise pixels' spectra being difficult to distinguish from respiration ones due to the spectral leakage, especially when the signal to noise ratio is low.

Moreover, considering complex environments as NICUs, the use of multiple camera views is important to obtain a proper coverage in all the infants possible positions for the detection of both respiratory flow and motion signals, called in this work respiration thermal signature. Furthermore, since neonates in NICUs are inside incubators and the incubators' Plexiglas walls are not transparent to long-wave infrared, thermal cameras should be positioned inside the incubator which is not feasible for large high resolution cameras. Even in case of babies being in an open bed, bulky cameras can obstruct caregivers' and parents' interaction with the infant, therefore, the use of smaller thermal solutions should be investigated.

Building on our previous work [20], this paper proposes a new algorithm based on an automatic ROI detection for respiration monitoring in multi-camera low-resolution thermal videos. We propose a data fusion on a pixel level and define three features that can be merged together to obtain a more accurate localization of the respiration pixels without relying on spectrum shape only, spatial averaging, or facial landmarks detection. This algorithm was tested on thermal videos recorded on seven infants in a real neonatal ward, reaching a total amount of 152 minutes using CI as gold standard. To our knowledge this is the first work showing results on such a large dataset of neonatal thermal recordings for respiration monitoring.

The rest of this paper is organized as follows. Section 2 explains the method developed, the setup used, and the dataset. Section 3 presents the results that are then discussed in Section 4. Section 5 highlights limitations and possible future research, and Section 6 contains the conclusion.

2. Materials and methods

2.1. Method

Three FLIR Lepton cameras were used to collect thermal videos of neonates in open bed. We chose to use multiple cameras to increase the camera coverage and have a good visualization of both respiration motion and flow considering different infants positions. Two cameras were positioned on the side of the open bed with a view that somewhat focuses on the head of the baby and the third one was mounted at the foot side of the bed registering the entire baby area. All the

cameras have a resolution of 60×80 pixels, for further details on the setup refer to Subsection 2.2 and Fig. 3.

The videos from the cameras are merged obtaining a single image-plane. Based on three features a core-pixel, presumed to contain a strong respiratory signature, is selected. This core-pixel is then combined with temporally highly correlating pixels to form the respiratory output signal. Consequently, this output uses both respiratory flow and motion present in all camera views. The Respiration Rate (RR) can then be estimated as the frequency corresponding to the peak of the spectrum. These steps are summarized in Fig. 1 and explained in more detail in the next sections.

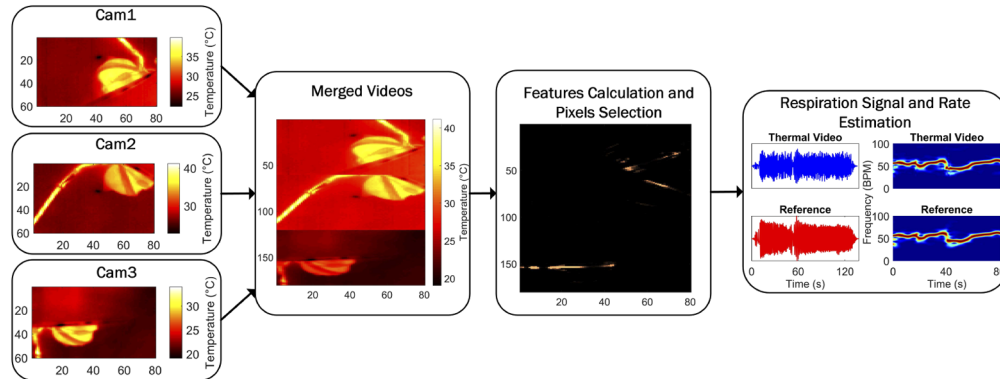


Fig. 1. Major processing steps summarized

Our algorithm will be benchmarked against Pereira's method [19], adapted to make it work with our hardware setup. The algorithm was developed and executed offline in MATLAB (MATLAB 2018a, The MathWorks Inc., Natick, MA, USA).

2.1.1. Preprocessing

The videos were recorded using MATLAB. However, due to the acquisition strategy the sampling rate of the videos was not uniform. Therefore, to obtain uniformly sampled videos, a 1D linear interpolation, i. e. using the standard MATLAB function *interp1*, was performed for each pixel's time domain signal resulting in three videos sampled at 9 Hz (the choice of the sampling frequency is not particularly critical and 9 Hz was chosen such that it was close to the average frame rate). At this point we considered two possible steps, one is treating the videos as three separate streams and then combine the RRs, the other one is combining the videos on a pixel level and process them as a single video. The first approach has some advantages based on the independence of each view, for example if a movement is visible only in one of the cameras the other two can potentially still deliver a good RR estimation. The second one, instead, rejects weaker respiration signatures, that can be obtained from unfavorable camera views, allowing to use only the overall best pixels. Most importantly however, the first method does not perform well when only one of the cameras detects a correct RR, and therefore, the second method has been implemented. The result, after merging the videos, is a single video with spatial dimensions equal to $M \times L$ with $M = 180$ and $L = 80$ pixels as shown in Fig. 1. It should be specified that we did not correct the single views' temperature values as visible in Fig. 1 because the absolute temperature values do not affect the processing since each pixel's time domain signal is used independently (the feature called thermal gradient is the only one that could be affected by differences in the temperature values as this could lead to a false edge-detection, however, the combination with the other features compensates for this effect).

The merged video was processed with a sliding window approach with a window size equal to 15 s and a slide of 1 s, the window size was chosen as a trade-off between accuracy and latency. A 15 seconds latency results in a resolution of 4 Breaths Per Minute (BPM), and the slide size results in an updated respiratory rate every second.

2.1.2. Automatic ROI detection

We based our respiration pixels selection strategy on different features that are used to find a core-pixel in the entire image, this pixel will be then used to find the other pixels that contain respiration. As the core-pixel is crucial to the result, the features are built to ensure a very strict selection. Each pixel is processed separately, let $x_i(nT_s)$ be each pixel's time domain signal in a 15 second window, with i being the index indicating the current pixel that can therefore go from 1 to 14400, i. e. the total number of pixels obtained after merging the three camera views. If we define the current 15 seconds time window as the j^{th} window then n is defined as $n = 0 + 9(j - 1), 1 + 9(j - 1), \dots, N + 9(j - 1)$ with N being equal to 135 samples, i. e. the samples in 15 seconds with a sampling time $T_s = 0.111$ s.

Three features were used to select the core-pixel in each window:

- **Pseudo-periodicity:** the first feature is similar to the one presented in our previous work [20], i. e. the height of the normalized spectrum peak that assumes the respiration signal to be pseudo-periodic and thus identifiable compared to noise. Each $x_i(nT_s)$ is filtered with a differential filter, the filtered version is named $x'_i(nT_s)$. This is transformed, after the application of Hanning window and zeropadded till $N_z = 120 \cdot N$, using a 1D Discrete Fourier Transform (DFT) obtaining $y'_i(f_k)$ with $k = 0, 1, \dots, \frac{N_z}{2} - 1$ and $f_k = \frac{k}{N_z T_s}$ Hz. and it is defined as

$$q_i = \frac{\max_{0 \leq f_k \leq \frac{(N_z/2-1)}{N_z T_s}} (|y'_i(f_k)|)}{\sqrt{\sum_{f_k=0}^{\frac{(N_z/2-1)}{N_z T_s}} |y'_i(f_k)|^2}}. \quad (1)$$

We obtain the first feature, then, by collecting all the q_i calculated for each pixel in a vector and by rearranging it in the matrix form, we obtain an $M \times L$ matrix called \mathbf{Q} , see example in Fig. 2. The feature defined in [20] resulted in not being accurate enough for the selection of a core-pixel in each j^{th} window without any previous knowledge, modification were therefore required. A constraint on the selection of the possible pixel, based on the proximity to the previously chosen pixels, was introduced in our previous work, which causes a dependency on the first selected core-pixel which is undesirable. Therefore, to avoid this dependency we introduced two new features.

- **Respiration rate spatial clusters:** the second feature comes from the consideration that pixels containing the respiration thermal signature are clustered in groups that present similar frequencies, i. e. the respiration rate. Therefore, for our second feature we select the frequency corresponding to the highest peak in the spectrum as:

$$rr_i = \frac{\arg \max_k (|y'_i(f_k)|)}{N_z T_s}. \quad (2)$$

The rr_i are arranged back into the image shape, called \mathbf{RR} with dimensions $M \times L$, on which a non-linear spatial filter with a 3×3 kernel is applied, as follows:

$$w_{m,l} = \frac{1}{9} \sum_{r=1}^3 \sum_{h=1}^3 \left(\frac{1}{\exp(70 |rr_{m,l} - rr_{r,h}| / rr_{m,l})} \right), \quad (3)$$

while r and h identify the kernel cell, m and l indicate the current central pixel in the entire $M \times L$ image. The constant 70 has been found to be not very critical and chosen such that it results in a weight equal to around 0.5 for a 1% relative error. Therefore, the resulting matrix will map the pixels having similar frequencies in their neighborhood and it is called \mathbf{W} , an example can be seen in Fig. 2.

- **Thermal gradient:** this feature assumes that respiration motion is visible in the thermal recordings and, considering that this is only visible if there is a thermal contrast, an edge map is built using the gradient as follows:

$$g_{m,l} = \begin{cases} 1 & \text{if } \sqrt{\left(\frac{\partial a_{m,l}}{\partial m}\right)^2 + \left(\frac{\partial a_{m,l}}{\partial l}\right)^2} > 1, \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

with \mathbf{A} being a thermal image in degrees Celsius representative of the current window j and evaluated as the average of all the thermal images in the window. A gradient of at least 1°C is considered, $a_{m,l}$ is an element of \mathbf{A} . The gradient operation is performed with the central difference method, using the standard MATLAB function *gradient*. The resulting matrix is called \mathbf{G} , Fig. 2 shows an example.

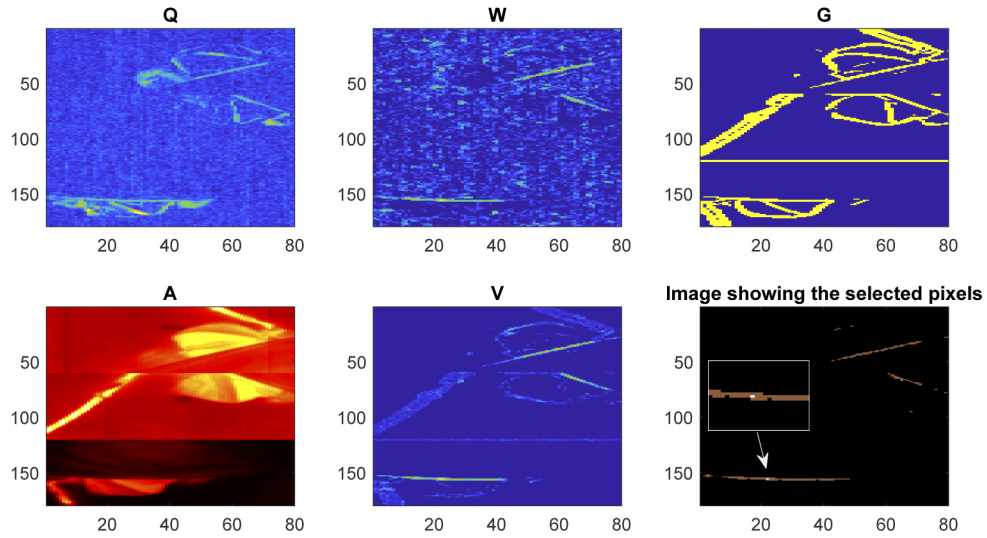


Fig. 2. On the top row an example of the three features used, on the bottom row a representative thermal image, the result of the multiplication, and a representation of all the pixels contained in \mathbf{P} . The core-pixel p_r is shown in white in the last figure.

Once the features are calculated, the matrices, \mathbf{Q} , \mathbf{W} , and \mathbf{G} are normalized to have values between 0 and 1 and then multiplied element-wise, obtaining a combination of the three features, called \mathbf{V} .

The core-pixel that is assumed to contain respiratory thermal signature is then selected as the pixel corresponding to the maximum of \mathbf{V} as:

$$p_r = \arg \max_i (v_i). \quad (5)$$

Once this pixel is selected, the other pixels containing respiration, both flow and motion, can be found based on the Pearson's correlation coefficient. More formally, the time domain signals,

$x_i(nT_s)$, are filtered at this point with a Butterworth bandpass filter from 30 *BPM* to 100 *BPM*, corresponding to the normal respiratory rate range in infants including also tachypnea cases [21], this filtered signal is called $\hat{x}_i(nT_s)$. The Pearson's correlation coefficient calculation on these DC-free signals as:

$$c_{p_r,s} = \frac{\sum_{t=1}^N (\hat{x}_{p_r}(t)) (\hat{x}_s(t))}{\left(\sum_{t=1}^N (\hat{x}_{p_r}(t))^2 \sum_{t=1}^N (\hat{x}_s(t))^2 \right)^{1/2}}. \quad (6)$$

The $c_{p_r,s}$ indicate the correlation between the chosen core-pixel p_r and a pixel in position s , that can vary from 1 to 14400, $\hat{x}_{p_r}(t)$ or $\hat{x}_s(t)$ are the filtered time domain signals of the p_r^{th} or s^{th} pixel, while t is an index that sweeps the time samples. In this work we considered pixels with an absolute correlation higher than 0.7, empirically chosen:

$$\mathbf{P} = s : |c_{p_r,s}| > 0.7. \quad (7)$$

\mathbf{P} consists of a set of pixels that are assumed to contain respiratory thermal signature. Figure 2 shows an example of all the features.

All pixels contained in \mathbf{P} are combined together with an average operation, after correcting for the sign of each time domain signal. The result is a single time domain signal representing the respiration signal detected in the j^{th} window. The signal is Hanning windowed and the spectrum is obtained through DFT. The RR is estimated as the frequency corresponding to the spectrum's peak for each window. A final time domain signal is obtained using an overlap-add procedure as explained in [22].

2.1.3. Benchmarking and reference

We benchmark our multi-camera and multi-feature ROI detection algorithm against Pereira's method [19] applied on all the three different view available in our dataset. A large spatial averaging is applied in Pereira's method, which was impossible to reproduce on our images due to the very different starting resolutions (our images have resolutions of 60×80 pixels while the images used in [19] have a resolution of 1024×768 pixels). Therefore, this step had to be skipped, as a consequence the quantization errors may be stronger in our dataset than in Pereira's set after down-scaling. Afterwards, a Hamming window is applied as done originally and at this point Pereira's method removes the DC component without further filtering. Directly using Pereira's method on our data, resulted in SQIs being higher for the noise pixels than the pixels showing the respiration signal. Consequently, the method failed as the wrong pixels were selected. This must be due to the two main differences in our setup: the lower spatial resolution, which prohibits further down-scaling, and the lower frame rate. Particularly, Pereira's method relies on spectrum features, as the one called F2, that depend on the frame rate. Therefore, we had to optimize Pereira's method to work on our data.

We were forced to apply a filter to attenuate the low part of the spectrum in order to get meaningful results. We used a differential filter as in our method, this filter does not change the essence of Pereira's method, so we feel it is a fair adaptation. DFT is applied to obtain the spectrum, and the magnitude of each spectrum is normalized for the maximum as done in Pereira's method. The SQI needs to be calculated at this point, in [19] three frequency bands were defined, a low-pass band below 0.1 *Hz*, a band-pass between 0.1 *Hz* and 3 *Hz* and a high-pass band above 3 *Hz*, and the SQI is calculated using four empirical spectrum features (i. e. the paper [19] mentions that F1 is the maximum amplitude in the high-pass band, F2 is the percentage of values in the high-pass band that are larger than an empirical threshold, F3 is the difference between the maximum amplitudes in band-pass and low-pass bands, and F4 is the ratio between the maximum amplitudes in the low-pass and band-pass bands). Pereira et al. empirically chose a threshold,

applied on the SQI, equal to 0.75 to eliminate the pixels that do not contain a respiratory signal. The three frequency bands so defined allow to use the method also for the detection of heart rate, whose typical frequencies are therefore included in the band-pass band.

In our case, heart rate detection is not one of the application's goals and also is not practically implementable due to our limited sampling frequency which would not respect the Nyquist-Shannon theorem (i. e. the sampling frequency should be at least two times the highest frequency of a band-limited signal, for heart rate detection in infants upper limit up to 5 Hz can be considered [23]). Therefore, we adapted the band to the same band we use in our application (low-pass below 0.5 Hz, band-pass between 0.5 Hz and 1.7 Hz, and high-pass higher than 1.7 Hz). The SQI has been calculated using the four features indicated in the original paper but some thresholds that were empirically defined had to be slightly adapted. Therefore, (i) the threshold for the calculation of the feature called F2 was increased from 0.1 to 0.2 and (ii) all the pixels having an SQI higher than 0.5 have been further processed for the RR calculation. We estimated the RR for each of the valid pixels and combined them using the median, obtaining a RR estimated for each window. The optimization of the SQI threshold for all the views and babies was the most problematic, as it resulted in a trade-off between inclusion of many noise pixels, and obtaining NaNs caused by absence of pixels meeting the criterion. We realized that using a higher threshold, e. g. 0.6 would be favorable resulting in a reduction of the error in some of the recordings, however, the use of a higher threshold could not prevent NaNs. Therefore, even though by tuning the thresholds for each single view and each single baby the results may get better, we decided to use the same thresholds for all the views which, in our opinion, is more realistic.

The RR was also estimated for the simultaneously collected CI signal in the same way, the CI signal was firstly filtered with a Butterworth bandpass filter from 30 BPM to 100 BPM. The RR so obtained constitutes our reference for both our method and Pereira's. Moving mean and median filters with 9 points were then applied in all cases.

To compare the results with the reference, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and a Percentage of correct values (PR) have been calculated. PR is defined as the percentage of estimations in which the absolute difference between an estimated RR and the reference RR is below 2 BPM, corresponding to the accuracy of a DFT with 4 BPM resolution. More formally:

$$PR = \frac{\sum_{j=1}^{N_w} (|rf_j - \tilde{rf}_j| < 2)}{N_w} \cdot 100, \quad (8)$$

where N_w is the number of windows, and rf_j and \tilde{rf}_j are respectively the CI respiration rate and the estimated respiration rate (ours or Pereira's) obtained in the j^{th} window. We analyzed the contribution of each of our features used for the detection of the core-pixel, and how the error varies when combining them. Moreover, to prove the dependency on window size of features based on the spectrum shape, we analyzed how the average MAE changes by keeping all the thresholds values and changing the window size to 8 s, for our method and for the benchmark one.

2.2. Experimental setup

The recordings were performed using three FLIR Lepton thermal cameras. These were Lepton 2.5, with a resolution of 60×80 , the cameras are sensitive in the Long-Wave Infrared (LWIR) range, specifically from 8 to 14 μm . The thermal sensitivity is around 50 mK and the average frame rate is 8.7 Hz. The sensors were connected to smart I/O modules (Pure Thermal 2 or Pure Thermal Mini) and these were connected to a laptop through USB. The acquisition was performed using MATLAB, the temperature readings were obtained by using the cameras with the factory default calibration. The cameras were positioned around the babies' open beds to cover all possible infant positions inside the bed, mounting arms as shown in Fig. 3 were used.

The two cameras pointing at the infant face from the two sides are called camera 1 and 2, while camera 3 is the camera on the foot side of the bed. We specify that camera 3 is actually a Lepton 3.5 which differs from the others only for the resolution, equal to 120×160 . We noticed the 60×80 resolution was sufficient in detecting the respiration rate and decided to use all the cameras at the same resolution. The images of the Lepton 3.5 were therefore down-scaled to 60×80 simulating the use of a third Lepton 2.5. All three cameras have a shutter which has been deactivated during the acquisition to avoid delays caused by the closing and opening of the shutter. The reference signal, CI, was obtained from the patient monitor (Philips MX700) sampled at 62.5 Hz . An artifact was used to synchronize the videos with the patient monitor.



Fig. 3. The three FLIR Lepton positioned around an infant's open bed, two with white housing and one with blue housing.

2.3. Dataset

The study was conducted in the neonatal ward of the Maxima Medical Center (MMC) in Veldhoven, The Netherlands and the study received a waiver from the Ethical Committee of the MMC. A written informed parental consent was obtained for all the infants included. The inclusion criteria were: (i) babies monitored using ECG electrodes for clinical purposes, (ii) clinically stable, (iii) in open bed. Both preterm and term babies were included in the study.

In total nine babies were included, two babies have been excluded from this analysis due to the blanket position that resulted in completely hiding the respiratory motion while the respiratory flow was also barely visible, Table 1 presents the information regarding the remaining seven

Table 1. Infants Data

Infant	Gender	Gestational Age (weeks+days)	Postnatal age (days)	Weight (g)	Sleeping Position	Duration (minutes)
1	F	26w 4d	59	2445	Supine	34.52
2	M	38w 5d	3	4140	Supine	9.24
3	M	34w 1d	16	2995	Supine	31.17
4	M	26w 3d	59	2750	Prone	11.10
5	M	39w	2	3670	Lateral	10.76
6	F	40w 1d	6	3420	Supine	44.70
7	M	40w 2d	1	3775	Lateral	10.17

infants. For each baby around three hours of videos were acquired, except for Infant 7 where only one hour was acquired. Since we expected problems from interfering motion (examples baby moving, parents or nurses handling, or soother presence) we selected only moments without patient handling and patient motion. The videos were, therefore, manually annotated, the useful moments can vary from around nine minutes to forty-five depending on the infant. One baby with respiratory support was included for feasibility purposes and corresponds to Infant 1 in Table 1.

3. Results

Table 2 provides the MAE, RMSE, and PR for our method and the benchmark applied on all the camera views. On the last row the average results considering all the infants is presented,

Table 2. Results showing MAE in BPM, RMSE in BPM, and percentage of correct estimation (PR) for our method and the benchmark one.

Infant	Ours			Benchmark								
	MAE	RMSE	PR	Cam 1			Cam 2			Cam 3		
	MAE	RMSE	PR	MAE	RMSE	PR	MAE	RMSE	PR	MAE	RMSE	PR
1	0.64	1.08	92.37	2.16	3.06	66.11	4.51	5.76	42.02	1.50	2.16	78.76
2	2.89	3.37	52.39	4.22	5.57	37.40	4.40	5.31	40.03	3.99	4.72	38.79
3	4.41	5.79	35.06	7.34	8.82	15.78	7.58	9.41	21.24	4.43	5.85	36.83
4	1.55	2.72	84.30	4.55	6.00	42.04	2.41	3.38	67.38	1.84	3.15	82.23
5	1.54	2.50	81.74	15.43	16.46	8.18	2.87	4.02	66.95	4.51	5.62	46.09
6	1.70	2.39	72.33	NaN	NaN	27.20	5.50	7.15	40.68	3.45	4.39	51.25
7	1.73	2.17	78.10	3.34	4.10	48.90	4.37	5.50	53.38	2.54	3.37	60.16
Average	2.07	2.86	70.90	NaN	NaN	35.09	4.52	5.79	47.38	3.18	4.18	56.30

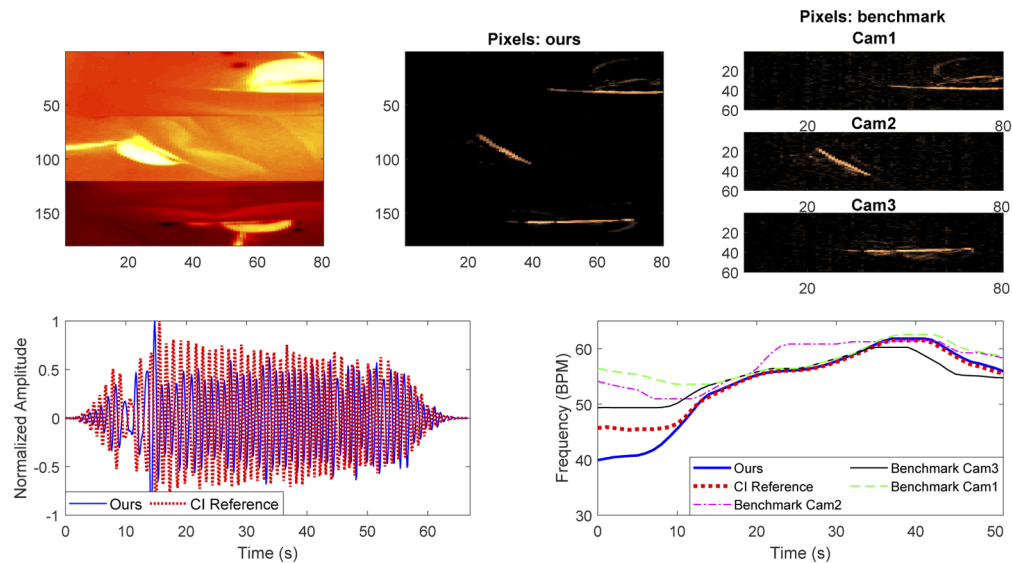


Fig. 4. Example of results obtained on Infant 1, on the top row the first image shows the merged thermal images from the three views, the second one and the third one show respectively the pixels selected for the respiration signal estimation with our proposed method and the benchmark method. On the bottom row instead the normalized time domain signals from the CI reference and our method and the RRs estimated in all the methods.

the best results for MAE, RMSE, and PR are indicated in bold. Our method yields an average MAE equal to 2.07 *BPM* and an average PR of 70.90%, while the benchmark obtains its best results on camera 3 with a MAE equal to 3.18 *BPM* and a PR of 56.30% (we also tested the benchmark on camera 3 at full resolution, i. e. 120×160 , the results were similar with a MAE equal to 3.39 *BPM*). Figure 4 shows an example of results obtained with the method proposed in this work and the benchmark method. A Correlation and Bland-Altman analysis was also performed for our method against the reference and displayed in Fig. 5. A mean bias of -0.55 *BPM* was obtained with the limits of agreement equal to -6.25 *BPM* and 5.14 *BPM*, and the Correlation plot shows the agreement between our estimated RRs and the reference ones obtaining a $\rho = 0.97$ ($p < 0.0001$).

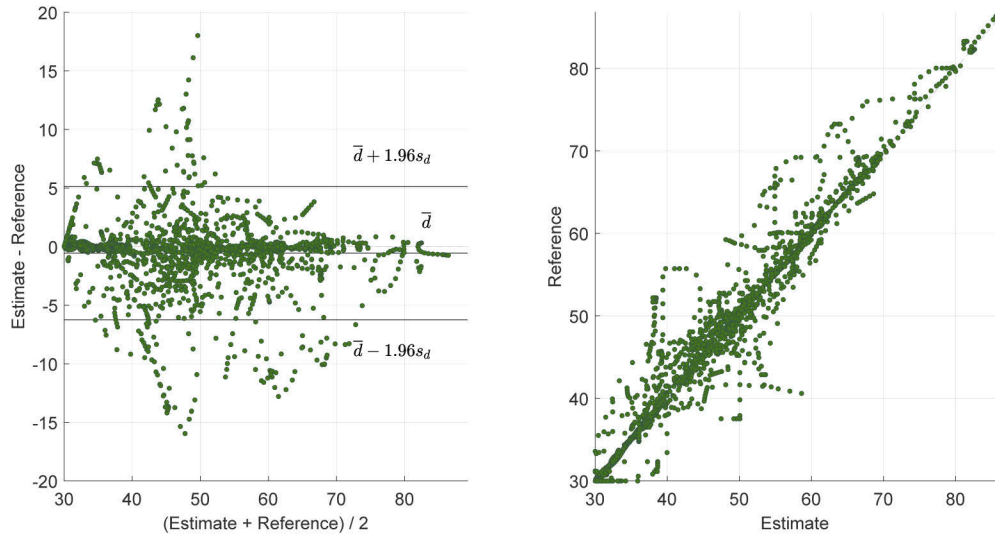


Fig. 5. Bland-Altman plot and Correlation plot. Estimate indicates the RR obtained with our method and Reference indicates the RR estimated using the CI.

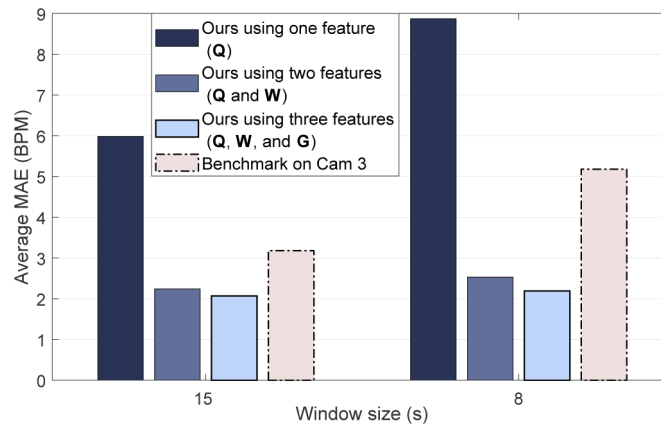


Fig. 6. Bar plot showing the average MAE obtained when using only the pseudo-periodicity feature (**Q**) or fusing it with the respiration rate clusters (**W**) and the thermal gradient (**G**) features. An analysis on how the average MAE changes due to the window size in our method and the benchmark is also included.

Fig. 6 shows how the individual features, used to select the-core pixel, influence the performance. The benchmark results on camera 3 are included in this figure to allow comparison. The results obtained with a window of 8 seconds are also shown in Fig. 6, the average MAE with this reduced window for all the babies in our method was 2.19 BPM, and 5.18 BPM for the benchmark applied on camera 3.

4. Discussion

In this work a new algorithm has been proposed to retrieve the respiration rate without facial landmark detection in low-resolution thermal recordings by combining pixels that highly correlate with a core-pixel that was carefully selected based on three new features. Moreover, we introduced the use of multiple parallel camera views, with data fusion on the pixel level, to enable respiration detection regardless of the momentary position of the babies. Our method was benchmarked against an adaptation of the method proposed by Pereira et al. [19].

The results obtained from around 152 minutes of infants thermal videos proved the method is able to correctly identify the pixels containing the respiration thermal signature yielding an average MAE of 2.07 BPM and around 71% of correct estimations, i. e. error below 2 BPM. The highest MAE was obtained in the case of Infant 3, this infant showed a Periodic Breathing (PB) pattern, which is a form of immature breathing typically present in newborns [24] causing fast changes in momentary RR. Our results were better than the benchmark, regardless of the camera view selected (the benchmark performed still worse than our method, even when the combined pixel data was provided as an input, and the benchmark algorithm again optimized for this different input). Camera 3 obtained the best benchmark result, this was expected since camera 1 and 2 may contain segments in which no flow or motion are visible in one of them due to the infant position. We further emphasize that to obtain the benchmark results, Pereira's method had to be further optimized in order to get meaningful results with our data.

The Bland-Altman plot in Fig. 5, shows the ability of our method to correctly estimate the RR, obtaining a mean bias of -0.55 BPM. The limits of agreement give an indication of the error spread $[-6.25; 5.14]$ BPM. Outliers are also present, these can be caused by synchronization problems that could be present in the data, considering that infants can have a fast changing RR small delays can cause high errors. Moreover, we should consider that the reference signal, i. e. the chest impedance, is also prone to errors, e. g. due to loose electrodes and baby-movements.

As visible from Fig. 6, using only the pseudo-periodicity feature was not reliable enough to select a core-pixel in each window, the use of a second and a third feature significantly improved the results. Moreover, our combination of three features proved to be more robust to spectrum changes due to a different window size yielding for a window of 8 seconds a MAE equal to 2.19 BPM compared to the significantly worsened results of the benchmark on camera 3: 5.18 BPM. We attribute this results to the use of respiration rate clusters and thermal gradient, that are not dependent on the spectrum shape.

Our low-resolution cameras were able to detect both respiratory motion and flow when the relative position of the baby and the camera were favorable. Babies in open bed are a particular population in neonatal wards and their habits and sleeping patters can significantly differ from babies in incubators. Infants in open bed are usually covered with blankets which can end up hiding the motion happening underneath, in two babies out of nine both flow and motion were not detectable, this was explained by the blanket position which ended up hiding the respiratory motion while the cameras did not have good position for the detection of the flow. A baby with respiratory support was also included in this study, Infant 1, and resulted in having the lowest MAE compared to all the other infants, this baby had a particularly quiet sleep compared to the others and the presence of the nasal cannulae introduced a thermal contrast region on the face itself increasing the number of pixels containing respiration motion.

5. Limitation of the study and future research

One of the most evident limitations is that we isolated around 152 minutes of useful moments from a total amount of 19 hours corresponding to the 13.3% of the time. This can be due to several reasons, firstly, events such as feeding, diaper changing, and nursing care took up to one hour with the baby being completely out of bed. Secondly, the algorithm is currently not able to cope with significant motion of the baby, or another person in the field of view. This implies that all segments including random motion of the baby, or nurses or parents hands in the field of view cannot be used. Since the babies included in this study were in open bed in a single-family room, where the presence of parents is highly encouraged to develop parent-infant bonding, segments where parents are comforting their baby (i. e. hands in the field of view) were quite frequent, as we shall discuss later. The occurrence of these moments should be considered to be significantly higher compared to the case of babies in incubators. Furthermore, also, the use of a soother is quite common in this population and causes a motion pattern that can result in being similar to respiration patterns but at a different frequency, these moments have, therefore, been excluded too.

A recent work studying a vast dataset of RGB video recordings of infants in incubators [4] showed after the annotation of 384.3 hours that in 54.9% of these the reference (CI) was poor, in the 11.5% the infant was out of the incubator, and in the 11.3% there were clinical interventions, and therefore, following our classification 22.3% would remain as useful data. CI is usually disturbed during baby motion and when there is a poor electrodes contact. Considering the different infant populations of the two studies we could attribute our reduced percentage of valid data to the use of the soother and to the increased parent presence. However, a more detailed annotation should be performed to analyze thoroughly the possible coverage obtainable with this type of setup and to investigate how to improve the algorithm or the setup for a more realistic clinical application.

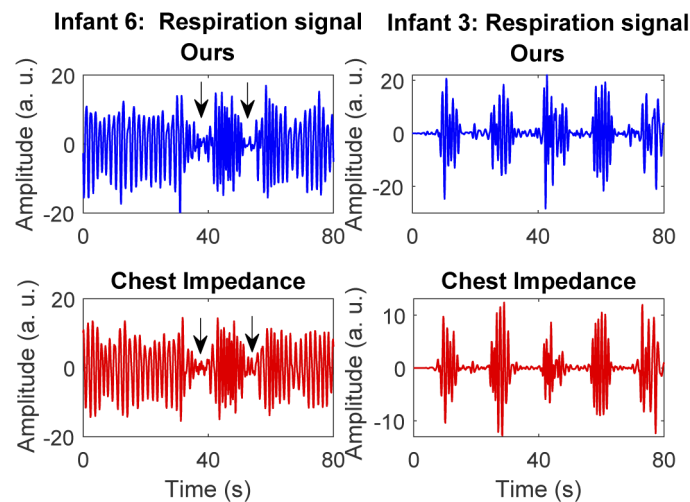


Fig. 7. Example of detectable cessation of breathing, the figures represents two segments from Infant 6 and 3, the first row contains the respiration signal obtained with our method and the second row shows the one obtained from the CI reference. Short apneic events (indicated with the arrows) were visually identified by looking at the thermal video and found also in the time domain signals, and PB pattern is clearly visible in the chest impedance and our signal in Infant 3.

This pilot study gave us good insights on the use of such technologies in a complex environment as a neonatal ward. Even though the focus of this paper was to verify the accuracy of our algorithm in detecting the respiration rate, there is evidence suggesting this type of setup could also be used to detect the occurrence of apneas and other type of patterns as shown in the examples in Fig. 7. Cessations of breathing were observed in Infant 6. These were clearly visible in the video thanks to the mattress changing temperature due to the respiratory flow. Infant 3 showed a typical PB pattern throughout most of the measurements. The detection of such events needs further analysis and should be part of the future development. Moreover, our algorithm currently detects the respiratory thermal signature, without distinguishing between motion and flow, this can be a limitation when aiming at apnea detection and should be further studied. Our method was tested in a neonatal ward setting including only patients in an open bed, the method, the setup, or the algorithm may need adaptation to work in a NICU setting with babies in incubator, due to the warm and humid environment inside incubators.

6. Conclusion

A novel approach for automatic respiration pixels detection in multi-camera thermal recordings was introduced. Our approach is based on the merging of the three camera views and the use of three features for the detection of the pixels containing respiration motion or flow. This type of approach has the advantage of being independent on the nose visibility required in approaches based on facial landmark detection and tracking. Our method was benchmarked against the method developed by Pereira et al., adapted to work with our recordings, and both were compared to the RRs obtained with the CI, on 152 minutes recordings acquired on seven infants. Our method yielded a percentage of correct estimation of around 71% compared to the benchmark best results of around 56%. Overall, this work shows results for the detection of the respiration rate in a quite large amount of data. Our algorithm obtained a MAE equal to 2.07 BPM by comparing our estimated RRs to the ones obtained from the CI reference, and using more features, not only dependent on the spectrum shape, guarantees robustness to window size changes.

Funding

Nederlandse Organisatie voor Wetenschappelijk Onderzoek (15345).

Acknowledgments

We would like to thank Astrid Osagiator and all the nurses working in the Medium Care Unit of MMC for their help and valuable contribution in our data collection.

Disclosures

The authors declare no conflicts of interest.

References

1. N. Rutter, "Applied physiology: the newborn skin," *Curr. Paediatr.* **13**(3), 226–230 (2003).
2. C. Battersby, S. Michaelides, M. Upton, and J. M. Rennie, "Term admissions to neonatal units in England: a role for transitional care? a retrospective cohort study," *BMJ open* **7**(5), e016050 (2017).
3. H. Lee, C. G. Rusin, D. E. Lake, M. T. Clark, L. Guin, T. J. Smoot, A. O. Paget-Brown, B. D. Vergales, J. Kattwinkel, J. R. Moorman, and J. B. Delos, "A new algorithm for detecting central apnea in neonates," *Physiol. Meas.* **33**(1), 1–17 (2012).
4. M. Villaruel, S. Chaichulee, J. Jorge, S. Davis, G. Green, C. Arteta, A. Zisserman, K. McCormick, P. Watkinson, and L. Tarassenko, "Non-contact physiological monitoring of preterm infants in the neonatal intensive care unit," *npj Digit. Med.* **2**(1), 128 (2019).
5. R. Janssen, W. Wang, A. Moço, and G. de Haan, "Video-based respiration monitoring with automatic region of interest detection," *Physiol. Meas.* **37**(1), 100–114 (2016).

6. M. Mercuri, I. R. Lorato, Y.-H. Liu, F. Wieringa, C. Van Hoof, and T. Torfs, "Vital-sign monitoring and spatial tracking of multiple people using a contactless radar-based sensor," *Nat. Electron.* **2**(6), 252–262 (2019).
7. J. D. Kim, W. H. Lee, Y. Lee, H. J. Lee, T. Cha, S. H. Kim, K.-M. Song, Y.-H. Lim, S. H. Cho, S. H. Cho, and H.-K. Park, "Non-contact respiration monitoring using impulse radio ultrawideband radar in neonates," *R. Soc. Open Sci.* **6**(6), 190149 (2019).
8. S. Nizami, A. Bekele, M. Hozayen, K. Greenwood, J. Harrold, and J. R. Green, "Comparing time and frequency domain estimation of neonatal respiratory rate using pressure-sensitive mats," in *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, (IEEE, 2017), pp. 239–244.
9. R. Joshi, B. Bierling, L. Feijs, C. van Pul, and P. Andriessen, "Monitoring the respiratory rate of preterm infants using an ultrathin film sensor embedded in the bedding: a comparative feasibility study," *Physiol. Meas.* **40**(4), 045003 (2019).
10. C. B. Pereira, X. Yu, M. Czaplik, V. Blazek, B. Venema, and S. Leonhardt, "Estimation of breathing rate in thermal imaging videos: a pilot study on healthy human subjects," *J. Clin. Monit. Comput.* **31**(6), 1241–1254 (2017).
11. L. J. Goldman, "Nasal airflow and thoracoabdominal motion in children using infrared thermographic video processing," *Pediatr. Pulmonol.* **47**(5), 476–486 (2012).
12. M. Usman, R. Evans, R. Saatchi, R. Kingshott, and H. Elphick, "Non-invasive respiration monitoring by thermal imaging to detect sleep apnoea," in *2019 32nd international congress and exhibition on condition monitoring and diagnostic engineering management (COMADEM)*, (2019).
13. A. K. Abbas, K. Heimann, K. Jergus, T. Orlikowsky, and S. Leonhardt, "Neonatal non-contact respiratory monitoring based on real-time infrared thermography," *BioMed. Eng. OnLine.* **10**(1), 93 (2011).
14. E. C. Eichenwald and A. A. P Committee on fetus and newborn, "Apnea of prematurity," *Pediatrics* **137**(1), e20153757 (2016).
15. C. B. Pereira, X. Yu, M. Czaplik, R. Rossaint, V. Blazek, and S. Leonhardt, "Remote monitoring of breathing dynamics using infrared thermography," *Biomed. Opt. Express* **6**(11), 4378–4394 (2015).
16. J. Fei and I. Pavlidis, "Thermistor at a distance: unobtrusive measurement of breathing," *IEEE Trans. Biomed. Eng.* **57**(4), 988–998 (2010).
17. Y. Cho, S. J. Julier, N. Marquardt, and N. Bianchi-Berthouze, "Robust tracking of respiratory rate in high-dynamic range scenes using mobile thermal imaging," *Biomed. Opt. Express* **8**(10), 4480–4503 (2017).
18. G. Scebbba, L. Tüshaus, and W. Karlen, "Multispectral camera fusion increases robustness of ROI detection for biosignal estimation with nearables in real-world scenarios," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (IEEE, 2018), pp. 5672–5675.
19. C. B. Pereira, X. Yu, T. Goos, I. Reiss, T. Orlikowsky, K. Heimann, B. Venema, V. Blazek, S. Leonhardt, and D. Teichmann, "Noncontact monitoring of respiratory rate in newborn infants using thermal imaging," *IEEE Trans. Biomed. Eng.* **66**(4), 1105–1114 (2019).
20. I. Lorato, T. Bakkes, S. Stuijk, M. Meftah, and G. De Haan, "Unobtrusive respiratory flow monitoring using a thermopile array: a feasibility study," *Appl. Sci.* **9**(12), 2449 (2019).
21. S. Reuter, C. Moser, and M. Baack, "Respiratory distress in the newborn," *Pediatr. review* **35**(10), 417–429 (2014).
22. G. De Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," *IEEE Trans. Biomed. Eng.* **60**(10), 2878–2886 (2013).
23. M. Villarroel, A. Guazzi, J. Jorge, S. Davis, P. Watkinson, G. Green, A. Shenvi, K. McCormick, and L. Tarassenko, "Continuous non-contact vital sign monitoring in neonatal intensive care unit," *Healthc. technology letters* **1**(3), 87–91 (2014).
24. M. Patel, M. Mohr, D. Lake, J. Delos, J. R. Moorman, R. A. Sinkin, J. Kattwinkel, and K. Fairchild, "Clinical associations with immature breathing in preterm infants: part 2—periodic breathing," *Pediatr. Res.* **80**(1), 28–34 (2016).